Neural Network Models for Spinal Implementation of Muscle Synergies

Yunqing Song[1], Masaya Hirashima[2,3] and Tomohiko Takei[1,4,5]*

[1] Graduate School of Medicine, Kyoto University, Kyoto, Japan, [2] Center for Information and Neural Networks (CiNet), Advanced ICT Research Institute, National Institute of Information and Communications Technology (NICT), Suita, Japan, [3] Graduate School of Frontier Biosciences, Osaka University, Suita, Japan, [4] The Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan, [5] Brain Science Institute, Tamagawa University, Machida, Japan

Muscle synergies have been proposed as functional modules to simplify the complexity of body motor control; however, their neural implementation is still unclear. Converging evidence suggests that output projections of the spinal premotor interneurons (PreM-INs) underlie the formation of muscle synergies, but they exhibit a substantial variation across neurons and exclude standard models assuming a small number of unitary "modules" in the spinal cord. Here we compared neural network models for muscle synergies to seek a biologically plausible model that reconciles previous clinical and electrophysiological findings. We examined three neural network models: one with random connections (non-synergy model), one with a small number of spinal synergies (simple synergy model), and one with a large number of spinal neurons representing muscle synergies with a certain variation (population synergy model). We found that the simple and population synergy models emulate the robustness of muscle synergies against cortical stroke observed in human stroke patients. Furthermore, the size of the spinal variation of the population synergy matched well with the variation in spinal PreM-INs recorded in monkeys. These results suggest that a spinal population with moderate variation is a biologically plausible model for the neural implementation of muscle synergies.

Keywords: muscle synergies, optimization, spinal interneurons, neural manifold, redundancy

INTRODUCTION

Our body is remarkably complex, yet we display a highly stable motor performance. For example, to reach for a coffee cup on a desk, there are an infinite number of patterns of muscle activity involved in extending the arm because multiple muscles span the same joint. Nevertheless, we show a highly stereotyped movement trajectory and agonist-antagonist muscle activity patterns (Morasso, 1981). Understanding how the central nervous system (CNS) coordinates the redundant musculoskeletal system is a central question of motor control.

Muscle synergies have been proposed as a solution to control redundant systems by coordinating a number of muscles with a smaller number of control modules, which are called muscle synergies (Tresch et al., 1999; Bizzi et al., 2002; d'Avella et al., 2003). This hypothesis is phenomenologically supported by experimental observations that a linear combination of basic patterns of muscle activity successfully reconstructs muscle activity during a wide range of behaviors, including reflex movements (Tresch et al., 1999; Cheung et al., 2005), postural tasks (Torres-Oviedo et al., 2006;

Ting and McKay, 2007), locomotion (Ivanenko et al., 2004; Krouchev et al., 2006; Dominici et al., 2011), reaching, and grasping (d'Avella et al., 2006; Overduin et al., 2008; Takei et al., 2017). However, there is still a heated debate regarding whether these experimental observations reflect a physiological basis of low-dimensional control in the CNS or an epiphenomenon based on task constraints and/or biomechanics (Tresch and Jarc, 2009; Kutch and Valero-Cuevas, 2012; Lillicrap and Scott, 2013; Hirashima and Oya, 2016).

Hirashima and Oya (2016) demonstrated that a neural network model that did not explicitly assume muscle synergies in the model could produce a synergy-like low-dimensional structure in muscle activity when the network was optimized to produce different combinations of elbow and shoulder torques while minimizing motor effort and motor error (Hirashima and Oya, 2016). This clearly shows that the mere fact that muscle synergies can be extracted from muscle activity is not enough to separate the presence or absence of underlying neural modules in the control system. Therefore, to examine the existence of muscle synergies, it is essential to identify the neural implementation of muscle synergies in the nervous system and develop a biologically plausible model.

Accumulating evidence from physiological and anatomical studies suggests that spinal premotor interneurons (PreM-INs) to motoneuron pools are the neural basis of muscle synergies in frogs (Kargo and Giszter, 2000, 2008; Hart and Giszter, 2010; Kargo et al., 2010), rodents (Levine et al., 2014), and primates (Takei and Seki, 2010, 2013; Takei et al., 2017). For the hindlimb control in frogs, it has been established that neural circuits for muscle synergies are implemented in the spinal cord (Kargo and Giszter, 2000, 2008; Kargo et al., 2010). Furthermore, output projections of PreM-INs identified by spike-triggered averaging showed a significant correlation with output patterns of extracted muscle synergies during wiping reflexes (Hart and Giszter, 2010). For the upper limb control in primates, our previous study demonstrated that cervical spinal PreM-INs have a divergent projection to multiple hand and arm motoneurons (Takei and Seki, 2010, 2013) and their spatial distribution corresponds to the spatial weight of muscle synergies extracted from muscle activity (Takei et al., 2017). These results suggest the contribution of spinal PreM-INs to the generation of muscle synergies. However, while the output projection of each PreM-IN corresponded to the muscle synergies, they also showed a substantial variation in the projection patterns across PreM-INs. Moreover, their temporal activation patterns were heterogeneous and were not clustered as muscle synergies at the individual neuron level. This finding clearly contradicts the assumption that each muscle synergy can be modeled as a unitary "module," where single or population neurons are synchronously activated by common inputs and act as fixed units. Such unitary modules are implicitly or explicitly assumed when applying linear decomposition methods such as non-negative matrix factorization (NNMF). Therefore, the development of neural models to explain how muscle synergies are implemented with divergent spinal PreM-INs has been awaited to reconcile physiological findings.

Here, we created neural network models for muscle synergies and compared their performance to explain a known

experimental phenomenon: the robustness of muscle synergies for cortical stroke (Cheung et al., 2009, 2012). In experiment 1, we examined the existence of muscle synergies by comparing two neural network models: one with random connections from the cortical layer to the muscle layer (non-synergy model) and the other with a small number of spinal synergies in the middle (simple synergy model). Then, in experiment 2, we sought a more biologically plausible model for muscle synergies by allowing a certain level of variation in spinal neurons that constitute muscle synergies (population synergy model). Our results demonstrate that (1) the simple and population synergy models emulate the robustness against cortical stroke that was observed in human stroke patients, and that (2) the population synergy model has a similar performance to the simple synergy model when the spinal variation was moderate, as in the variations of spinal PreM-IN recorded in non-human primates in our previous study (Takei et al., 2017). These results suggest that the population synergy model with a moderate spinal variation is a biologically plausible model for neural implementation of muscle synergy to achieve robust motor control.

MATERIALS AND METHODS

Neural Network Models

We compared three types of neural network models for neural implementation of muscle synergy: (1) non-synergy model, (2) simple synergy model, and (3) population synergy model.

Non-synergy Model

The non-synergy model was the same as the neural network model in previous studies (Hirashima and Nozaki, 2012; Hirashima and Oya, 2016). Briefly, a feedforward neural network was used to convert the desired torque (input layer) into the actual torque (output layer) through an intermediate layer composed of 1,000 cortical neurons and eight muscles (**Figure 1A**). Each cortical neuron received the desired torque vector (τ) from the input layer with synaptic weights (W). Then, the cortical neurons projected to the muscles with innervation weights (Z). The innervation weights from each cortical neuron to muscles were established so that Z_i ($i = 1$– 1,000) were uniformly distributed on the surface of a sphere in an 8-dimensional space, with a radius of 0.002 ($=2/n$). The eight muscles included two shoulder flexors (outer and inner shoulder flexors: SFo and SFi, respectively), two shoulder extensors (outer and inner shoulder extensors: SEo and SEi, respectively), an elbow flexor and extensor (EF and EE, respectively), and a biarticular flexor and extensor (BiF and BiE, respectively). The mechanical effects of the muscles (mechanical direction vectors, MD vectors) were in line with those of the previous model (M) (Hirashima and Nozaki, 2012; Hirashima and Oya, 2016). Muscle activity (a) was constrained to be non-negative by replacing negative muscle activity with zero, which made the model non-linear ($a \geq 0$). The total output of the network (T) was expressed as the vector sum of the mechanical output of all muscles. The

synaptic weight from the input layer to the cortical neurons was initially selected from a standard normal distribution with a zero mean and unitary standard deviation. Then, it was modified by the error feedback-with-decay algorithm (Hirashima and Nozaki, 2012; Hirashima and Oya, 2016) (see section "Training Procedure").

Simple Synergy Model

In the simple synergy model, we added a spinal synergy layer between the cortical and muscle layers (**Figure 1B**). Other than the addition of the spinal synergy layer, the simple synergy model was identical to the non-synergy model. The cortical output to the spinal synergies (X) was similarly defined as the non-synergy model (Z), but the target number was different. To define the spinal synergies, we extracted muscle synergies from muscle activations produced by the optimized non-synergy model. After optimizing the non-synergy model in the learning simulation of 40,000 trials, we applied a non-negative matrix factorization (NNMF) to the muscle activation patterns during the last 100 trials of the training session. To determine the number of muscle synergies, we performed a four-fold cross validation of NNMF for the different numbers of muscle synergies (1–7), and plotted the averaged variance accounted for (VAF) as a function of the number of muscle synergies (**Figure 1E**). VAF was calculated as VAF = 1 – SSE/SST, where SSE is the sum of the squares of residual errors, and SST is the sum of the square differences between each data point and the overall mean. Similar to Hirashima and Oya (2016), we successfully extracted four muscle synergies from their non-synergy network model based on the criterion that the VAF exceeds 0.90 (Hirashima and Oya, 2016). To maintain consistency in the muscle synergy extraction, we set the number of synergies to four for the following NNMF analyses. The synergy weights (Sy1–4) of the non-synergy model were used to define the output of the four spinal synergies (V) in the simple synergy model. In this regard, the synergies explicitly defined in the simple synergy model were identical to the muscle synergies that were originally extracted from the non-synergy model, which are referred to as original synergies. This procedure is particularly critical when testing the robustness of muscle synergies because the previous study demonstrated that different synergies could be extracted when NNMF is executed from different initial values (Hirashima and Oya, 2016). By applying the muscle synergies extracted from the non-synergy model to the other model, we controlled the initial synergy structure of both models to be the same.

Population Synergy Model

In experiment 2, we further tested another synergy model, in which there were 100 spinal neurons ($n = 100$). Initially, each group of 25 spinal neurons was provided one of the synergy weights (Sy1–4) for their outputs (V). Then, by adding a different proportion of Gaussian noise to the synergy weights (V), we systematically created population synergy models with different spinal variation levels. The output weights of each spinal neuron (V_i, $i = 1$–100) was obtained by mixing the original synergy weights (Sy$_k$, $k = 1$–4) with Gaussian noise (ω), which was

normalized to the norm to be unitary, with a certain proportion (ρ):

$$V_i = (1 - \rho)Sy + \rho\omega$$

By changing the proportion of noise (ρ) from 0 to 1, we systematically investigated the synergy models which had different spinal variations, replicating a variation of the spinal PreM-INs recorded in monkeys performing a precision grip task (Takei et al., 2017). The silhouette value was calculated to evaluate the clustering of the spinal neurons with regard to their similarity to the original synergies (Sy1–4) (Rousseeuw, 1987).

Isometric Torque Production Task

We trained each network to perform isometric torque production tasks using a two-joint system (shoulder and elbow) (Herter et al., 2007). The target torque combination was chosen from eight possible torque combinations [shoulder flexion (SF), shoulder and elbow flexion (SF + EF), elbow flexion (EF), shoulder extension and elbow flexion (SE + EF), shoulder extension (SE), shoulder and elbow extension (SE + EE), elbow extension (EE), and shoulder flexion and elbow extension (SF + EE)] with the norm of 1 Nm.
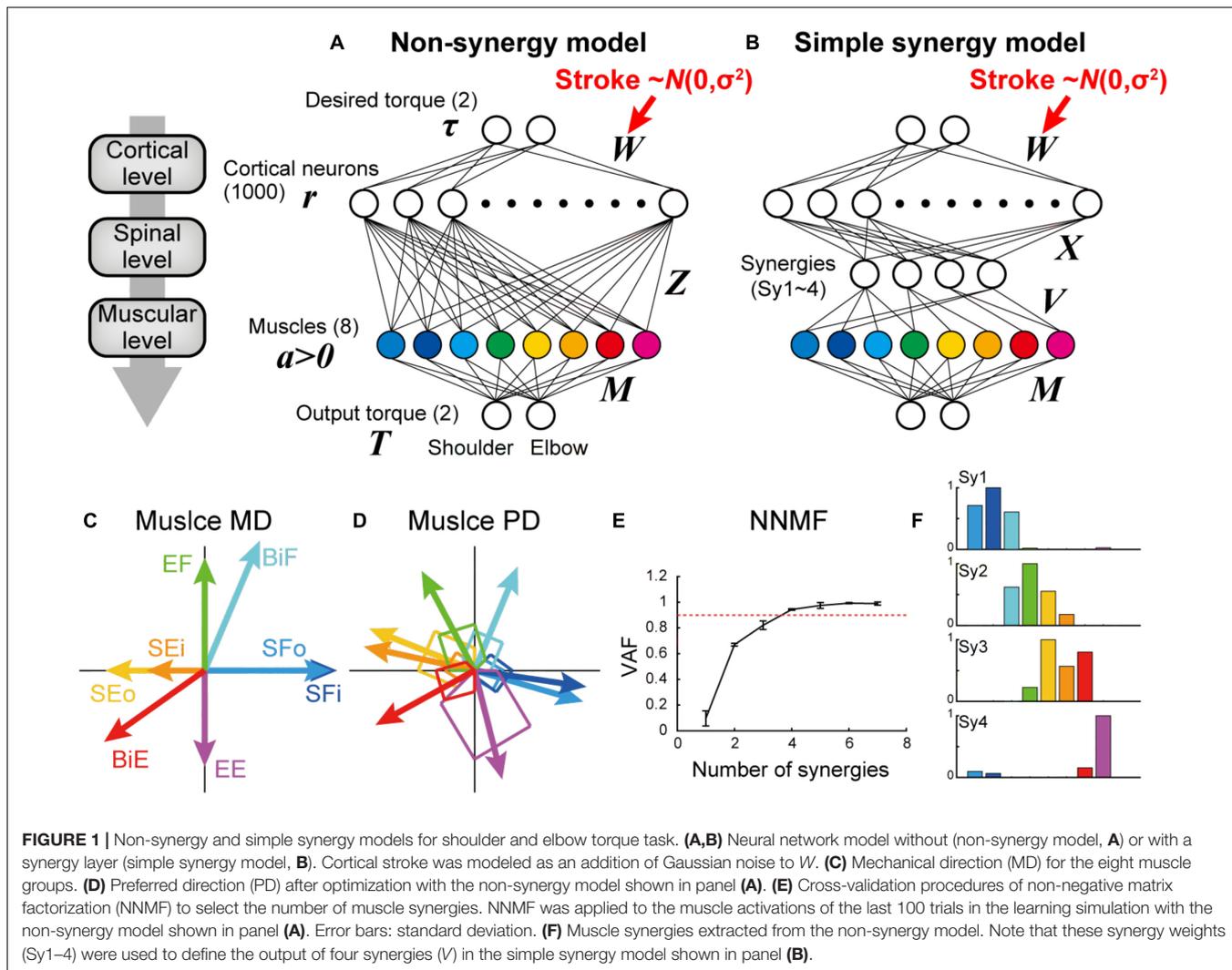
Training Procedure

The synaptic weight from the input layer to the cortical neurons (W) was modified using the error feedback-with-decay algorithm (Hirashima and Nozaki, 2012; Hirashima and Oya, 2016):

$$\Delta W_{ij} = -\alpha \frac{\partial J_e}{\partial W_{ij}} - \beta W_{ij}$$

where, α is the learning rate ($\alpha = 20$) and J_e is the error cost, as calculated by the error vector ($e = T - \tau$) between the output torque and the desired torque: $J_e = 1/2 e^T e$. The second term indicates that the changes in synaptic weight due to synaptic memory decay in each step are proportional to the current synaptic weight W_{ij}. The decay rate β was set to 1.0×10^{-4}, which is much smaller than the learning rate ($\alpha = 20$). By randomly presenting one of the eight target torque combinations in 40,000 trials and modifying the synaptic weight (W) after each trial, the network was trained to produce the appropriate output torque. The step-by-step update of W for the non-synergy model can be written as:

$$W(t + 1) = W(t) + \Delta W(t)$$
$$= W(t) + \left(-\alpha \frac{\partial J_e}{\partial W(t)} - \beta W(t)\right)$$
$$= W(t) - \alpha Z^T M^T (MZW(t) - I) \tau(t) \tau(t)^T$$
$$-\beta W(t) \qquad (1)$$

This update procedure has been mathematically proven to reach an optimal solution for linear models (Hirashima and Nozaki, 2012). To incorporate the non-linearity of our model, where muscle activity is constrained to be non-negative ($a \geq 0$), we modified M by replacing the MD vector with $[0,0]^T$ if the

FIGURE 1 | Non-synergy and simple synergy models for shoulder and elbow torque task. **(A,B)** Neural network model without (non-synergy model, **A**) or with a synergy layer (simple synergy model, **B**). Cortical stroke was modeled as an addition of Gaussian noise to *W*. **(C)** Mechanical direction (MD) for the eight muscle groups. **(D)** Preferred direction (PD) after optimization with the non-synergy model shown in panel **(A)**. **(E)** Cross-validation procedures of non-negative matrix factorization (NNMF) to select the number of muscle synergies. NNMF was applied to the muscle activations of the last 100 trials in the learning simulation with the non-synergy model shown in panel **(A)**. Error bars: standard deviation. **(F)** Muscle synergies extracted from the non-synergy model. Note that these synergy weights (Sy1–4) were used to define the output of four synergies (*V*) in the simple synergy model shown in panel **(B)**.

muscle activity generated at each step $[a = ZW(t)\tau(t)]$ was negative ($a < 0$). With the modified MD vector, $\tilde{M}$, the update equation [Eq. (1)] can be rewritten as:

$$W(t+1) = W(t) - \alpha Z^T \tilde{M}^T \left(\tilde{M} Z W(t) - I \right) \tau(t) \tau(t)^T$$
$$- \beta W(t) \qquad (2)$$

This procedure eliminates the effect of negative muscle activity by disabling the mechanical output of the muscles. Our previous numerical simulation demonstrated that our training procedure can make the non-linear system reach a solution qualitatively similar to the optimal solution derived by the linear model. This was demonstrated by the fact that the synaptic weight matrix (*W*) is aligned orthogonally to the mechanical property matrix (*M*) of the muscle, which is an essential property of the optimal solution (i.e., pseudo-inverse of *M*) (Hirashima and Nozaki, 2012). For the synergy model (simple or population synergy model), *Z* in Eq. (1) and (2) is replaced with *VX*.

Stroke Model

To simulate the situation of cortical stroke, we added Gaussian noise ($\sim N[0, \sigma^2]$) to the cortical synaptic weights (Stroke, **Figures 1A,B**, **6A,B**), since previous human imaging studies showed that the long-scale entropy of MEG signal increased in perilesional tissues in stroke patients (Kielar et al., 2016). We systematically changed the stroke size (σ) from 0 to 10 standard deviations (SD) of the original synaptic weights (*W*). We sampled 200 different stroke cases for each stroke size ($\sigma = 0$–10) and evaluated their task performance.

Evaluation of Task Performance and Muscle Synergies

To evaluate the stroke effect on task performance and muscle synergies, we used 96 target torque combinations (32 uniformly distributed torque directions × 3 amplitudes [0.5, 1, and 2 Nm]) instead of the eight trained torque combinations. We quantified two types of performance errors, directional error and amplitude error, as a function of stroke size. The directional error was defined as the angle between the target torque vector and the

torque vector output by the model. The amplitude error was defined as the difference between the amplitude of the target torque and the model output. Both errors were calculated for each stroke size, and the absolute values were averaged across the 200 stroke cases.

To evaluate the consistency of muscle activation patterns, we quantified the shift in the preferred direction (ΔPD). PD was defined as the direction in which the muscle is maximally active and it was calculated by the summation of the target torque vectors weighted by muscle activity. ΔPD was defined as the absolute shift in PD from no stroke cases ($\sigma = 0$).

To evaluate the robustness of muscle synergies in cortical stroke, we quantified two different measures: VAF by the original muscle synergies and similarity of muscle synergies to the original muscle synergies. For these evaluations, we used muscle activations during the same 96 target torque combinations. To evaluate the VAF by the original muscle synergies (Sy1–4), we applied NNMF for muscle activation without updating the synergy weights (Sy1–4). We then calculated the VAF between the observed and reconstructed muscle activity. This measure indicates how much variance in muscle activity can be explained by the original muscle synergies (Sy1–4). Note that the synergy weights (V) for the simple and population synergy models were derived from the original muscle synergies extracted from the non-synergy model (see section "Neural Network Models"). Therefore, we expect that the VAF by the original muscle synergies will be the same for all synergy models under the intact ($\sigma = 0$). and no spinal variation ($\rho = 0$) conditions. We also compared the similarity of muscle synergies extracted in each stroke case with the original muscle synergies. We applied NNMF to muscle activity to extract muscle synergies for each stroke case. For this extraction, we did not use n-fold cross validation, and the number of muscle synergies was fixed at four to allow for a consistent comparison. Then, we calculated the dot product between the extracted synergies and the original synergies. We used the max dot product to determine the similarity of each muscle synergy and averaged the values across the four synergies. This measurement represents the consistency of the muscle synergies between different stroke conditions.

Statistical Analyses

To test the significance of the effect of different neural network models on performance measures (directional error, amplitude error, ΔPD, VAF by the original synergies, and similarity with the original synergies), we used the bootstrap method. For the two groups of parameter samples ($n = 200$), we first calculated the original difference in the mean of two populations (Δmean). Then, we pooled the two populations, resampled two populations of the same size with replacement, and calculated the Δmean. This process was repeated 1,000 times to obtain a baseline distribution of the Δmean. We set the significance limits of this distribution to 0.23 (= 2.5/number of stroke conditions) and 99.77 (= 100 − 2.5/number of stroke conditions) percentile that matched a significance level of $p < 0.05$, with Bonferroni's correction. All simulations
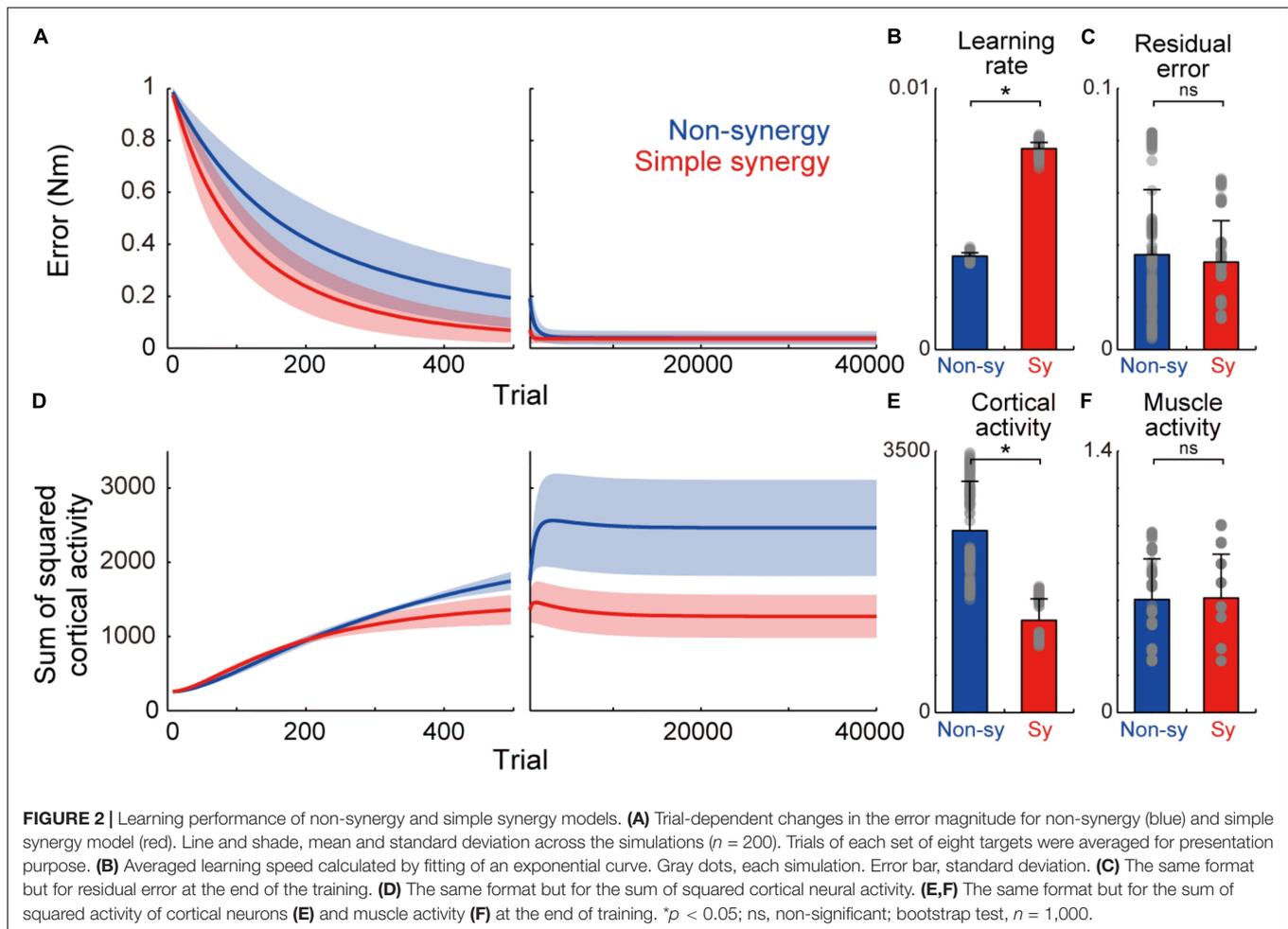
and analytical procedures were performed using MATLAB (MathWorks, RRID:SCR_001622).

RESULTS

Non-synergy and Simple Synergy Models Have Comparable Training Ability

In experiment 1, to test the existence of muscle synergies in the nervous system, we compared the task performance of the neural network model with and without muscle synergies (simple synergy vs. non-synergy models). In the non-synergy model, cortical neurons had random direct connections to the muscles (Z). Cortical activation patterns (W) were optimized to achieve the model to output the desired shoulder and elbow torques while minimizing the sum of squares of cortical activity (**Figure 1A**). As Hirashima and Oya (2016) demonstrated previously, the non-synergy model reproduced a shift in the preferred direction of muscles (**Figure 1D**) relative to their mechanical directions (**Figure 1C**), and the muscle activation patterns were successfully reconstructed with a linear combination of four muscle synergies (VAF $\geq$ 0.9, **Figures 1E,F**). We refer to the synergies extracted from the non-synergy model as the original synergies (Sy1–4, **Figure 1F**). In the simple synergy model, we added four spinal synergies between the cortical and muscle layers (**Figure 1B**). We set the output weight of the synergies (V) to be the same as the synergy weights extracted from the non-synergy model (Sy1–4, **Figure 1F**). Other than the addition of the synergy layer, the simple synergy model was the same as the non-synergy model.

First, we compared the learning performance of the non-synergy and simple synergy models. Both models were trained to produce eight combinations of shoulder and elbow torques while minimizing the sum of squared cortical activity (i.e., minimizing error cost and motor cost). **Figure 2A** shows the trial-dependent changes in the error magnitude averaged across the eight target conditions. After a 40,000-trial training, we found that learning converged in both models, and they reached similar residual errors (**Figure 2C**). However, during the training, the learning speeds differed. At the initial phase of the training (**Figure 2A**, left), the learning curve was steeper in the simple synergy model than in the non-synergy model, and the learning rate, fitted by an exponential curve, was higher in the simple synergy model (**Figure 2B**, $p < 0.05$, bootstrap test, $n = 1,000$). This indicates that learning progressed faster in the simple synergy model than in the non-synergy model. This result is consistent with the previous study, which examined the learning performance of similar network models (Hagio and Kouzaki, 2018). Furthermore, we also tested trial-dependent changes in the sum of squared cortical neural activity (**Figure 2D**). We found that after training, cortical activity was less for the simple synergy model than for the non-synergy model (**Figure 2E**). **Figure 3A** compares the distribution of activity size of cortical neurons in both models. This shows that the distribution is narrower in the simple synergy model than in the non-synergy model, and that the activity size is generally reduced rather than changed in a specific group of neurons. Interestingly, despite the prominent difference in cortical activity, the sum of squared muscle activity was the same for both models
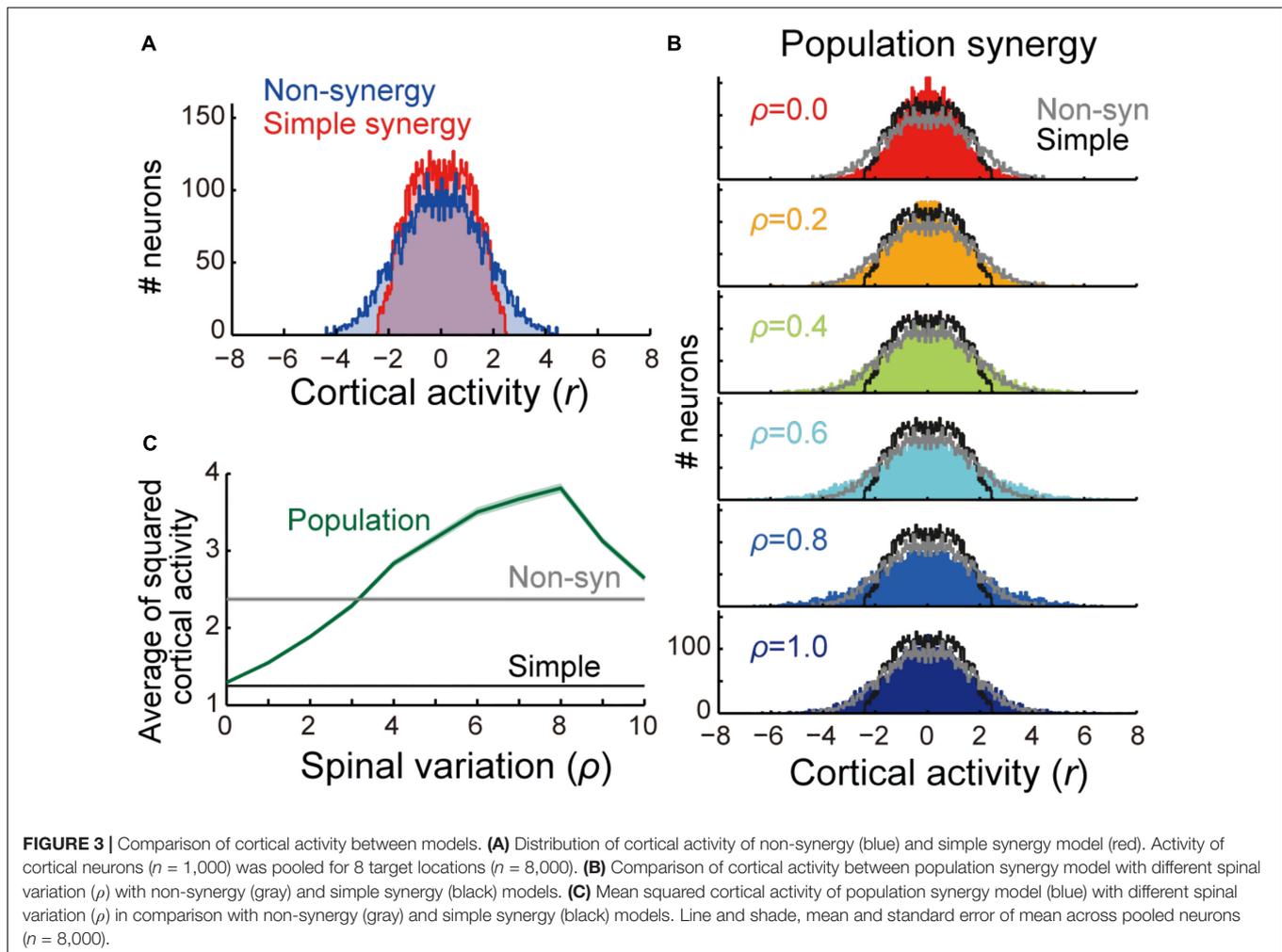
FIGURE 2 | Learning performance of non-synergy and simple synergy models. **(A)** Trial-dependent changes in the error magnitude for non-synergy (blue) and simple synergy model (red). Line and shade, mean and standard deviation across the simulations (n = 200). Trials of each set of eight targets were averaged for presentation purpose. **(B)** Averaged learning speed calculated by fitting of an exponential curve. Gray dots, each simulation. Error bar, standard deviation. **(C)** The same format but for residual error at the end of the training. **(D)** The same format but for the sum of squared cortical neural activity. **(E,F)** The same format but for the sum of squared activity of cortical neurons **(E)** and muscle activity **(F)** at the end of training. *$p < 0.05$; ns, non-significant; bootstrap test, $n = 1,000$.

(**Figure 2F**). These results indicate that the simple synergy model showed faster convergence of learning and smaller neural activity, although both models exhibited similar task performance.

Simple Synergy Model Exhibits a Higher Robustness Against Cortical Stroke Than Non-synergy Model

We compared the robustness of non-synergy and simple synergy models against cortical stroke. To simulate the situation of cortical stroke, we added random noise to the cortical connections (W) of different sizes (σ = 0–10 SD). We tested 200 different stroke cases for each stroke size. **Figure 4** shows three examples of torque output by the non-synergy model (**Figure 4A**) and simple synergy model (**Figure 4B**) in no-stroke and in two different sizes of stroke (σ = 1 and 2 SD). Both models showed increasing instability as the stroke size increased. However, systematic differences exist in the distribution of the errors. While the errors of the non-synergy model were distributed almost evenly in all directions (i.e., circular shape), the errors of the simple synergy model were distributed more ovally, and the main axis was along the target direction. For example, for the shoulder flexion target (the rightmost target) with σ = 2SD, the error was

distributed horizontally in the simple synergy model (**Figure 4B**); however, it was distributed evenly in the non-synergy model (**Figure 4A**). To examine the details of the task performance, we evaluated two types of errors: directional error and amplitude error. The directional error is the angle between the target torque vector and the torque vector output by the model, while the amplitude error is the difference between the amplitude of the target torque and the output of the model. **Figures 4C,D** show an increase in the directional and amplitude errors as a function of stroke size. The figures show that whereas the amplitude error of both the non-synergy and simple synergy models increased in a similar manner, the simple synergy model produced a smaller size of directional error ($p < 0.05$, bootstrap test, $n = 1,000$) although the effect size was relatively small (Cohen's d was ranged from 0.15 to 0.94). This result suggests that the simple synergy model retains the ability to coordinate muscle activity even when cortical activity is disturbed, and robustly generates joint torques in the correct direction.

Next, we compared the robustness of muscle activation patterns between non-synergy and simple synergy models. **Figures 5A,B** illustrate muscle activation for each target direction with non-synergy and simple synergy models in no-stroke and two different sizes of stroke cases (σ = 3 and 6 SD). While muscle
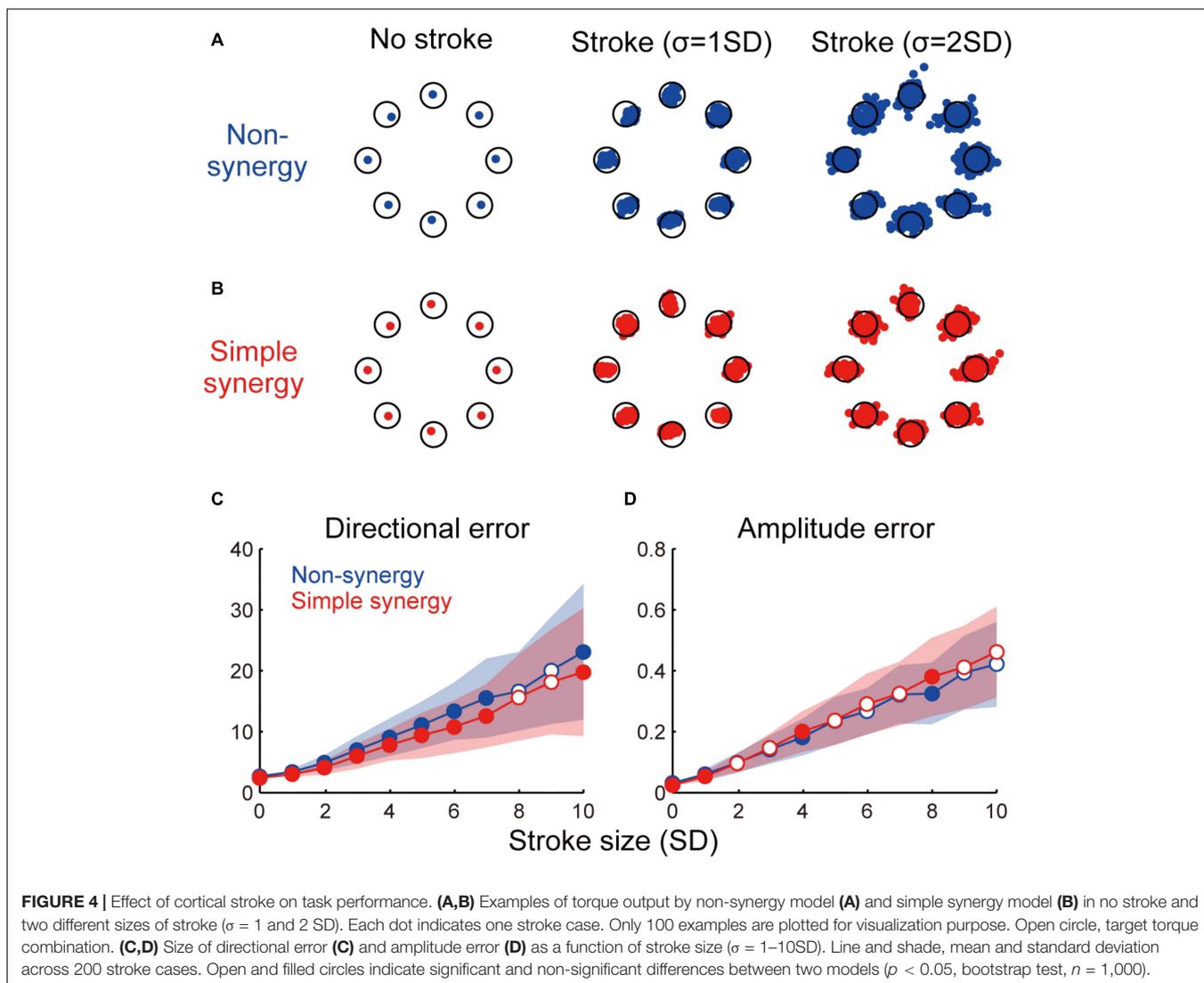
FIGURE 3 | Comparison of cortical activity between models. **(A)** Distribution of cortical activity of non-synergy (blue) and simple synergy model (red). Activity of cortical neurons ($n = 1,000$) was pooled for 8 target locations ($n = 8,000$). **(B)** Comparison of cortical activity between population synergy model with different spinal variation (ρ) with non-synergy (gray) and simple synergy (black) models. **(C)** Mean squared cortical activity of population synergy model (blue) with different spinal variation (ρ) in comparison with non-synergy (gray) and simple synergy (black) models. Line and shade, mean and standard error of mean across pooled neurons ($n = 8,000$).

activation of the non-synergy model was severely disturbed by stroke, the simple synergy model expressed more consistent muscle activation. The shift in PD (ΔPD) was quantified as the absolute difference of the PD from no-stroke cases ($\sigma = 0$). **Figure 5C** shows that ΔPD was significantly smaller for the simple synergy model than for the non-synergy model ($p < 0.05$, bootstrap test, $n = 1,000$). This result indicates that the simple synergy model is more robust for cortical stroke to generate consistent muscle activation.

We further compared the robustness of muscle synergies between non-synergy and simple synergy models. To quantify the robustness of muscle synergies, we compared two measures: (1) how much variance of affected muscle activity can be accounted for by original synergies (VAF by original synergies) and (2) how similar the synergies extracted from the affected muscle activity was to the original synergies (similarity with original synergies). **Figures 5D,E** demonstrate that while both measures steeply decrease in the non-synergy model, they remain high in the simple synergy model (≥ 0.90, **Figure 5D**, red dotted line), even when the stroke size increased to 10 SD. The difference in the effect of stroke size on each model was significant for both measures ($p < 0.05$, bootstrap test, $n = 1,000$). These results

indicate that the simple synergy model more robustly generates consistent muscle synergies under stroke conditions.

Population Synergy Model With a Moderate Spinal Variation Exhibits a Comparable Robustness to Simple Synergy Model

We examined how the synergies could be organized in the spinal layer. The simplest model is that in which each synergy is represented by a single neuron or unitary "module" (simple synergy model, **Figures 1B**, **6A**). However, our previous physiological study showed that a muscle field of individual spinal PreM-INs did not completely match muscle synergies but showed substantial variation in spatial and temporal activation patterns (Takei et al., 2017). Therefore, a more plausible model can be developed in such a way that a synergy is represented by a population of neurons with some variation (population synergy model, **Figure 6B**). To test this scenario, in experiment 2, we compared the robustness of these synergy models against the introduction of cortical stroke. First, we systematically changed the size of the variation of synergies by changing the proportion

FIGURE 4 | Effect of cortical stroke on task performance. **(A,B)** Examples of torque output by non-synergy model **(A)** and simple synergy model **(B)** in no stroke and two different sizes of stroke (σ = 1 and 2 SD). Each dot indicates one stroke case. Only 100 examples are plotted for visualization purpose. Open circle, target torque combination. **(C,D)** Size of directional error **(C)** and amplitude error **(D)** as a function of stroke size (σ = 1–10SD). Line and shade, mean and standard deviation across 200 stroke cases. Open and filled circles indicate significant and non-significant differences between two models ($p < 0.05$, bootstrap test, $n = 1,000$).

(ρ) of Gaussian noise to the synergy weights (V). We found that spatial clustering of the output weights of each spinal neuron based on the similarity to the original muscle synergies decreased as the variation level ρ increased (**Figure 6C**). The level of clustering was evaluated using silhouette values (0.93, 0.81, and 0.56 for $\rho = 0.2$, 0.3, and 0.4, respectively) (Rousseeuw, 1987). Importantly, our previous electrophysiological study demonstrated that the similarity of output projection of spinal PreM-INs to extracted muscle synergies showed a spatial clustering and their silhouette values ranged from 0.79 to 0.87 (Figure 2A of Takei et al., 2017). This silhouette value was comparable to that of the population synergy model, with $\rho = 0.3$. After network training procedures, all population synergy models showed a similar level of task performance, with residual errors ranging from 0.017 to 0.083, comparable to the performance of the non-synergy and the simple synergy models (**Figure 2C**, 95% interval: 0.006 – 0.083 and 0.012 – 0.064, respectively). The amount of cortical activity generally increased as spinal variability increased, and it exceeded the non-synergy model when spinal variability was greater than 0.3 which was the physiologically observed level ($\rho > 0.3$, **Figures 3B,C**).

Then, we compared the robustness of task performance and muscle synergies in cortical stroke between the simple and population synergy models. **Figures 7A–E** shows the change in task performance, ΔPD, and muscle synergies as an effect of cortical stroke. As a reference, we also plot the results of the non-synergy model (**Figures 7A–E**, gray rectangles). In general, results of the population synergy models spanned from the results of the simple synergy model to those of the non-synergy model. For example, the directional error of the population synergy model was similar to the simple synergy model when the spinal variation was smaller ($\rho = 0.0–0.2$, **Figures 7A–E**, red lines), while it was similar to the non-synergy model when the variation was larger ($\rho = 0.8–1.0$, blue lines). This gradual change was observed for directional error, ΔPD, and two muscle synergy measures ($p < 0.05$, bootstrap test, $n = 1,000$), but not for the amplitude error (**Figure 7B**). Critically, the robustness of the population synergy models remained almost at the same
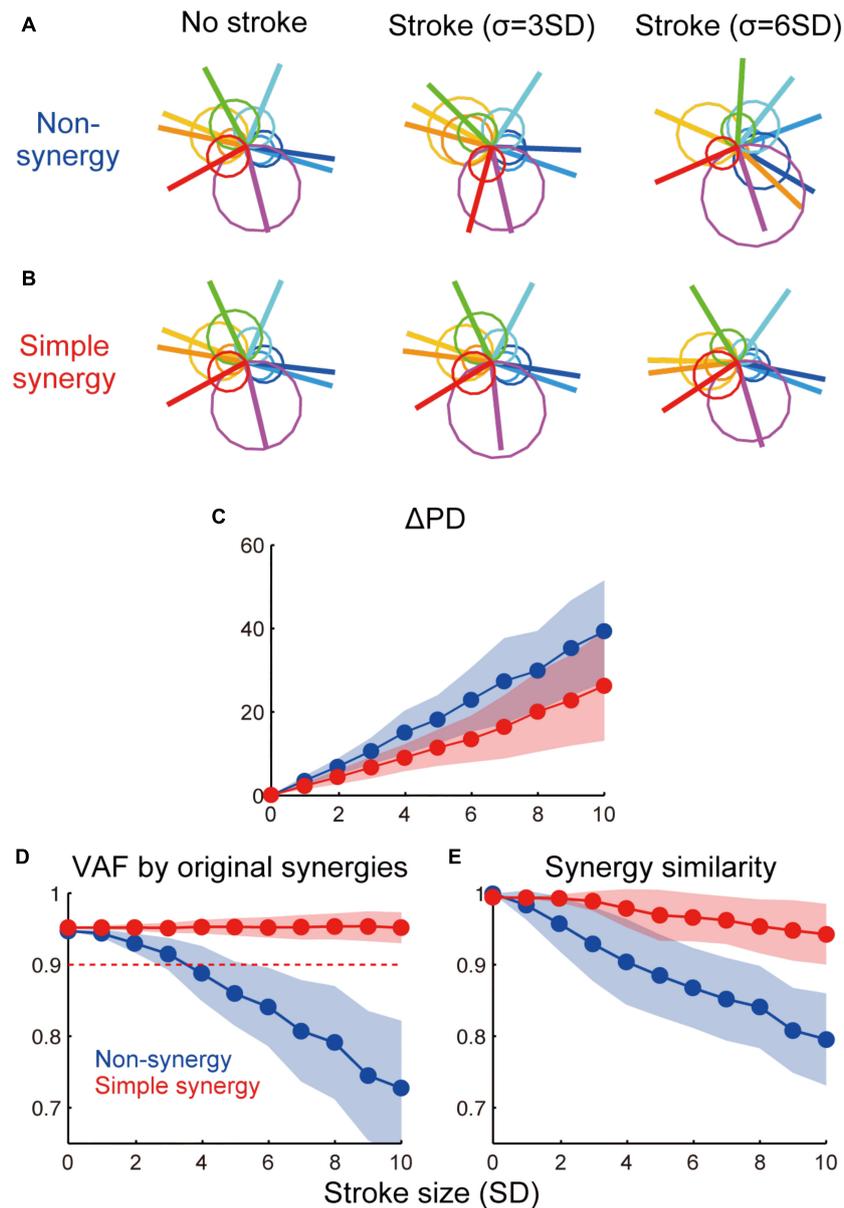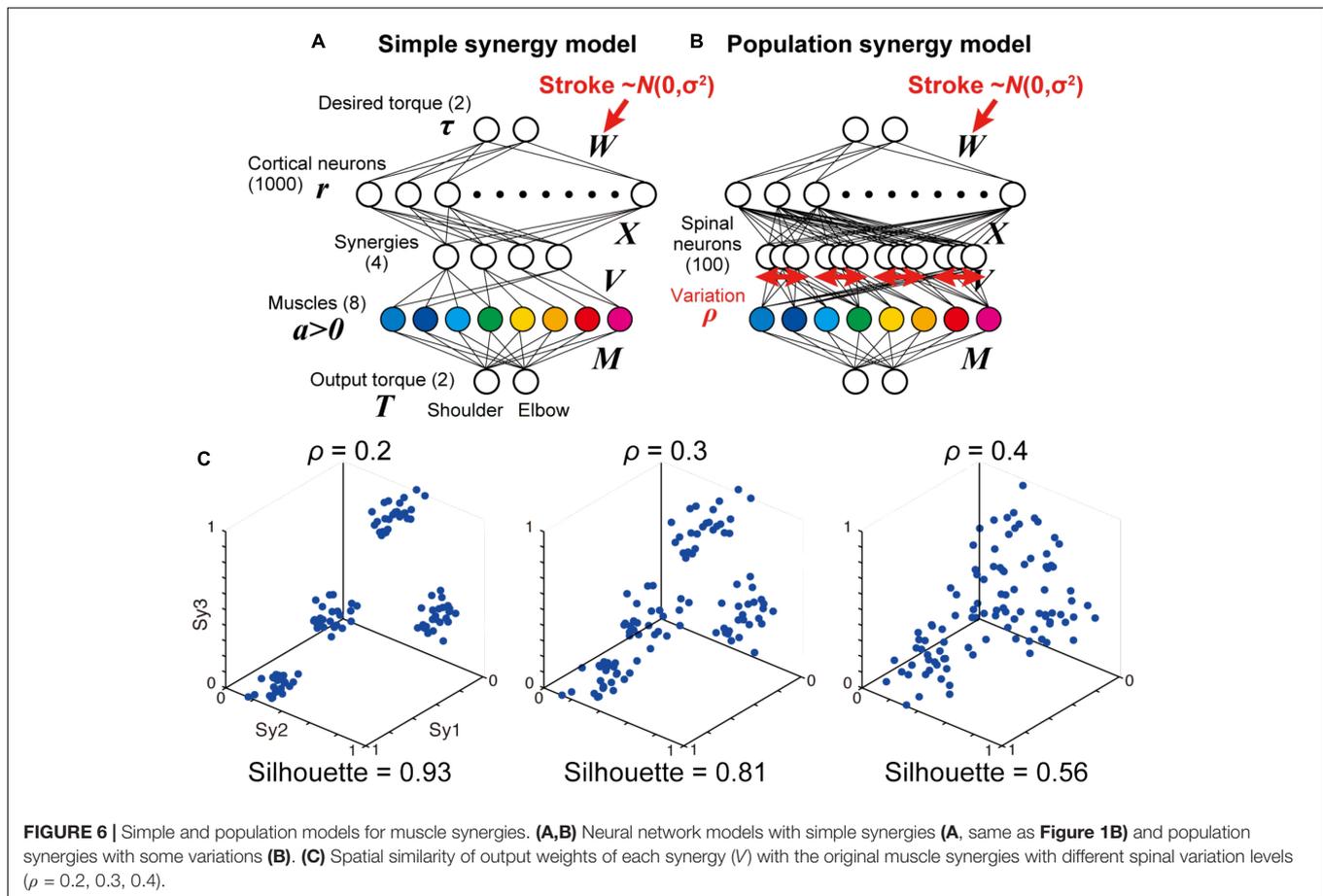
FIGURE 5 | Effect of cortical stroke on muscle synergies. **(A,B)** Examples of muscle activation with non-synergy model **(A)** and simple synergy model **(B)** in no-stroke and two different sizes of stroke cases ($\sigma = 3$ and 6 SD). Circles indicate muscle activation for each torque direction and lines indicate a PD of muscles. **(C)** Averaged shift of PD (ΔPD) from no stroke ($\sigma = 0$) condition. **(D)** VAF by original muscle synergies extracted from non-synergy model and used for simple synergy model **(Figure 1F)**. **(E)** Similarity of muscle synergies extracted in each stroke case to the original muscle synergies. The format is the same as **Figures 4C,D**.

level as the simple synergy model until the spinal variation level was increased to 0.3 (**Figures 7C–E**, red-yellow), and then these measures steeply decreased when the variation increased further (green-blue). These results indicate that the population synergy model with moderate variation exhibited robustness against cortical stroke comparable to that of the simple muscle synergy model. These results suggest that the muscle synergies are not necessarily implemented as unitary "modules," but they can be implemented as the population of spinal neurons with a moderate variation.

DISCUSSION

In this study, we compared different neural network models with a focus on their robustness to cortical stroke. In experiment 1, we compared two types of neural network models: one with random connections (non-synergy model) and the other with a smaller number of spinal synergies (simple synergy model). After optimization, we found that both models achieved comparable task performance and similar muscle synergies, confirming the prediction of a previous study (Hirashima and Oya, 2016).
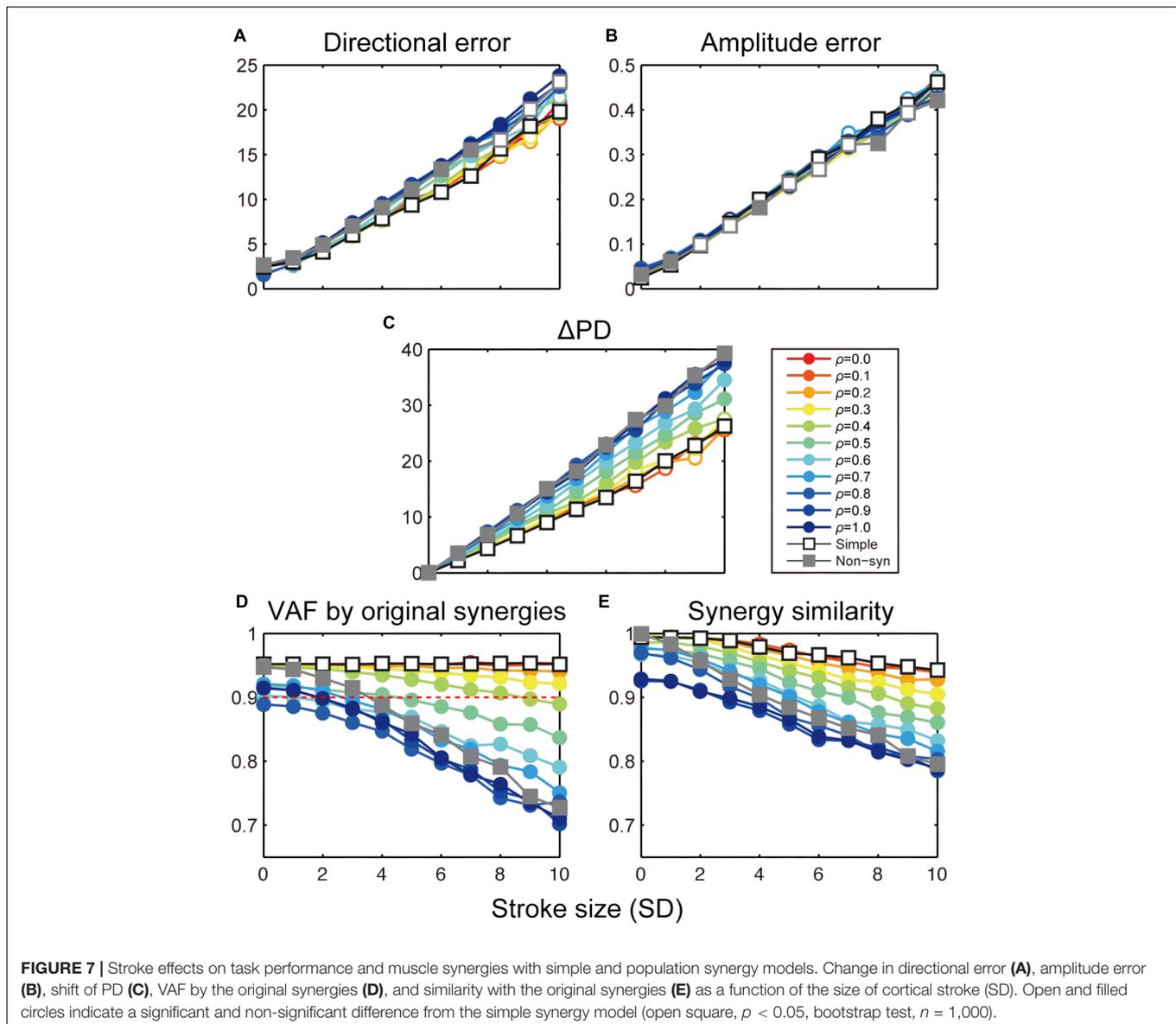
FIGURE 6 | Simple and population models for muscle synergies. **(A,B)** Neural network models with simple synergies (**A**, same as **Figure 1B**) and population synergies with some variations **(B)**. **(C)** Spatial similarity of output weights of each synergy (*V*) with the original muscle synergies with different spinal variation levels (ρ = 0.2, 0.3, 0.4).

Interestingly, despite the similar task performance after the training, the simple synergy model had higher learning rates, consistent with previous findings (Hagio and Kouzaki, 2018), and smaller cortical activity (**Figures 2A–F**). This indicates that although the output performances were similar, the simple synergy model achieved more efficient muscle control with less neural activity cost than the non-synergy model.

Then, we tested the robustness of the model performance against cortical stroke, which was modeled as the addition of noise to the cortical layer. The results demonstrated that the simple synergy model exhibited (1) smaller directional error in torque production and (2) higher consistency of muscle synergies with the original muscle synergies. This robustness of muscle synergies was consistent with observations in human stroke patients (Cheung et al., 2009, 2012). compared muscle synergies extracted from the unaffected and affected arms of stroke patients with moderate-to-severe unilateral ischemic lesions in the frontal motor areas (Cheung et al., 2009). Their results demonstrated that most of the patients showed muscle synergies that were strikingly similar between the unaffected and affected arms despite differences in motor performance between the arms. The similarity evaluated by a dot product between muscle synergies from the arms ranged from 0.80 to 0.95 with an average of 0.90. Our results showed that the similarity of muscle synergies with the original synergies was maintained at >0.90, even

with a severe stroke in the simple synergy model (**Figure 5D**, red dotted line). The consistency of our results with previous findings suggests that the existence of a synergy layer is an essential factor in explaining the robustness of muscle synergies in stroke conditions.

We further explored a biologically plausible model for the neural implementation of muscle synergies. Our previous study demonstrated that spinal PreM-INs have divergent output projections to multiple hand and arm motoneurons (Takei and Seki, 2010, 2013) and their spatial distribution corresponds to the spatial weight of muscle synergies (Takei et al., 2017). However, although the output projection of PreM-INs corresponded to muscle synergies as a population, there was a substantial variation in the projection patterns across individual PreM-INs. Moreover, their temporal activation patterns were heterogeneous and were not clustered as muscle synergies at the individual neuron level. From these observations, we hypothesized that each muscle synergy is not represented by a unitary "module," in which single or population neurons are synchronously activated by common inputs, but by a population of PreM-INs that have variable output projections and are activated independently (Takei et al., 2017). Consistent with this hypothesis, the present results demonstrated that the population synergy models with a spinal variation exhibited comparable robustness of muscle synergies with the simple synergy model against cortical stroke

FIGURE 7 | Stroke effects on task performance and muscle synergies with simple and population synergy models. Change in directional error **(A)**, amplitude error **(B)**, shift of PD **(C)**, VAF by the original synergies **(D)**, and similarity with the original synergies **(E)** as a function of the size of cortical stroke (SD). Open and filled circles indicate a significant and non-significant difference from the simple synergy model (open square, $p < 0.05$, bootstrap test, $n = 1,000$).

(Cheung et al., 2009, 2012). The similarity of muscle synergy to the original synergies remained high (>0.90, **Figure 7D**, red dotted line) even when the spinal variation level was increased up to $\rho \leq 0.3$ (**Figures 7D,E**, red-yellow). Importantly, the variation of spinal PreM-INs reported in our previous study (Takei et al., 2017) corresponded to $\rho = 0.3$, according to the silhouette value calculated from the cluster analysis for spinal PreM-INs (**Figure 2A**; Takei et al., 2017). This variation level was within the range where the present models preserved the robustness of muscle synergies against cortical stroke. These findings suggest that the population of spinal neurons with a moderate variation is a biologically plausible model for the neural implementation of muscle synergies. Note that previous studies on muscle synergy modeling also utilize a variation or noise addition, but their purpose was to test the robustness and degradation of the model performance (Kargo et al., 2010; Ausborn et al., 2019), which

corresponds to cortical stroke in the present study. On the other hand, we added a spinal variation to change the synergy structure so that each spinal neuron has different output projections and receives input from cortical neurons independently, which allows spinal neurons to be activated asynchronously. This is qualitatively different from the conventional "module" models, which assume that the population of neurons in the same module are synchronously activated by common input. Our finding showed that the population synergy model with a moderate spinal variation showed high robustness and similarity to the monkey physiological data.

This finding has several implications for the low-dimensional control of the limb and muscles. First, muscle synergies do not necessarily reflect the reduction of the degrees of freedom (DOFs) by a smaller number of unitary modules. Previously, muscle synergy has often been modeled as a unitary module, where single

or population neurons are synchronously activated by common inputs and act as fixed units, and has been interpreted as a simplified strategy for motor control. However, the spinal PreM-IN showed substantial variation between neurons to contradict such a simple view of modular control (Takei et al., 2017). One possible explanation for the variation of spinal PreM-INs is that these spinal PreM-INs represent another muscle synergy that was not recruited in the task tested (a precision grip task). Indeed, for the control of primate dexterous hand movements, it is suggested that the higher-order control is working with the lower-order control which is usually detected as synergies (Santello et al., 1998; Yan et al., 2020). It is of interest to further examine how the spinal variability identified here can contribute to the higher-order control for dexterous hand movements. Another explanation is that the muscle synergies are represented by a population of PreM-INs with a certain variability of input and output connections. The present simulation results demonstrated that the population of spinal neurons which have a moderate variation has an ability to robustly control muscle activity to achieve motor coordination in task space. Despite the apparent higher dimensionality of the population synergy model, to achieve the motor coordination in task space the variability of spinal neurons will necessarily be averaged in the output or may effectively be in a null space of the output-potent dimension. To exploit the variations in the population synergy, more selective or sparse recruitment of subsets the population synergy layer may be required and then added amplifications of this output may be needed in a real spinal system. It is noteworthy that a similar constraint of neural variability to task space has been identified as neural manifolds in the population activity of motor cortices (Kaufman et al., 2014; Elsayed et al., 2016; Gallego et al., 2017). It is of interest to investigate how the cortical manifold and spinal synergy space interact and are hierarchically organized to regulate the dimensionality of motor control.

Our model of spinal population with a moderate variation in muscle synergy also provides implications for motor learning and development of motor coordination. Previous studies have demonstrated that motor tasks compatible with the original muscle synergy are learned faster than tasks incompatible with muscle synergies (Berger et al., 2013). Interestingly, the results also showed that although the learning rate was slow during the incompatible task, learning progressed to explore new synergies. During this exploration process, the variation in spinal neurons may be involved in changing the muscle coordination patterns. It is also interesting to understand how muscle synergies are obtained during development. In the sensory system, it is known that the synaptic connection in primary sensory cortices increases after birth till it reaches a maximum, then prunes and decreases as neural selectivity increases (Huttenlocher et al., 1982). In contrast, in the motor system it has been suggested that

muscle synergy is generally conserved from early development to adulthood in a variety of motor animals (Dominici et al., 2011; Giszter, 2015; Yang et al., 2019) and that many aspects of afferent controls and premotor patterning may be independent of early stage functional activity (Haverkamp, 1986; Mendelson and Frank, 1991). However, at the same time, these studies also demonstrated that there are some modifications of muscle synergies over the course of development, including increased inter-subject variability (Yang et al., 2019) and recruitment of additional synergies (Dominici et al., 2011). It would be of interest to investigate how the variation of spinal PreM-INs shown in this study relates to these developmental and learning modifications of muscle synergies.

In conclusion, our network simulation showed that muscle synergies could be implemented as a population of PreM-INs with a moderate variation rather than unitary "modules." This view provides new insight into understanding the mechanism and functional relevance of how the CNS controls the redundant motor system.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

TT designed the study. YS and TT carried out the experiments and analyzed the data. YS, MH, and TT wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ausborn, J., Shevtsova, N. A., Caggiano, V., Danner, S. M., and Rybak, I. A. (2019). Computational modeling of brainstem circuits controlling locomotor frequency and gait. *Elife* 8:e43587. doi: 10.7554/eLife.43587

Berger, D. J., Gentner, R., Edmunds, T., Pai, D. K., and d'Avella, A. (2013). Differences in adaptation rates after virtual surgeries provide direct evidence

for modularity. *J. Neurosci.* 33, 12384–12394. doi: 10.1523/JNEUROSCI.0122-13.2013

Bizzi, E., d'Avella, A., Saltiel, P., and Tresch, M. (2002). Modular organization of spinal motor systems. *Neuroscientist* 8, 437–442. doi: 10.1177/107385802236969

Cheung, V. C., Piron, L., Agostini, M., Silvoni, S., Turolla, A., and Bizzi, E. (2009). Stability of muscle synergies for voluntary actions after cortical stroke

in humans. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19563–19568. doi: 10.1073/pnas. 0910114106

Cheung, V. C., Turolla, A., Agostini, M., Silvoni, S., Bennis, C., Kasi, P., et al. (2012). Muscle synergy patterns as physiological markers of motor cortical damage. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14652–14656. doi: 10.1073/pnas.1212 056109

Cheung, V. C. K., d'Avella, A., Tresch, M. C., and Bizzi, E. (2005). Central and sensory contributions to the activation and organization of muscle synergies during natural motor behaviors. *J. Neurosci.* 25, 6419–6434. doi: 10.1523/ JNEUROSCI.4904-04.2005

d'Avella, A., Portone, A., Fernandez, L., and Lacquaniti, F. (2006). Control of fast-reaching movements by muscle synergy combinations. *J. Neurosci.* 26, 7791–7810. doi: 10.1523/JNEUROSCI.0830-06.2006

d'Avella, A., Saltiel, P., and Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. *Nat. Neurosci.* 6, 300–308. doi: 10.1038/nn1010

Dominici, N., Ivanenko, Y. P., Cappellini, G., d'Avella, A., Mondì, V., Cicchese, M., et al. (2011). Locomotor primitives in newborn babies and their development. *Science* 334, 997–999. doi: 10.1126/science.1210617

Elsayed, G. F., Lara, A. H., Kaufman, M. T., Churchland, M. M., and Cunningham, J. P. (2016). Reorganization between preparatory and movement population responses in motor cortex. *Nat. Commun.* 7:13239. doi: 10.1038/ncomms13239

Gallego, J. A., Perich, M. G., Miller, L. E., and Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron* 94, 978–984. doi: 10.1016/j.neuron.2017. 05.025

Giszter, S. F. (2015). Motor primitives-new data and future questions. *Curr. Opin. Neurobiol.* 33, 156–165. doi: 10.1016/j.conb.2015.04.004

Hagio, S., and Kouzaki, M. (2018). Modularity speeds up motor learning by overcoming mechanical bias in musculoskeletal geometry. *J. R. Soc. Inter.* 15:20180249. doi: 10.1098/rsif.2018.0249

Hart, C. B., and Giszter, S. F. (2010). A neural basis for motor primitives in the spinal cord. *J. Neurosci.* 30, 1322–1336. doi: 10.1523/JNEUROSCI.5894-08. 2010

Haverkamp, L. J. (1986). Anatomical and physiological development of the Xenopus embryonic motor system in the absence of neural activity. *J. Neurosci.* 6, 1338–1348. doi: 10.1523/JNEUROSCI.06-05-01338.1986

Herter, T. M., Kurtzer, I., Cabel, D. W., Haunts, K. A., and Scott, S. H. (2007). Characterization of torque-related activity in primary motor cortex during a multijoint postural task. *J. Neurophysiol.* 97, 2887–2899. doi: 10.1152/jn.00757. 2006

Hirashima, M., and Nozaki, D. (2012). Learning with slight forgetting optimizes sensorimotor transformation in redundant motor systems. *PLoS Comput. Biol.* 8:e1002590. doi: 10.1371/journal.pcbi.1002590

Hirashima, M., and Oya, T. (2016). How does the brain solve muscle redundancy? Filling the gap between optimization and muscle synergy hypotheses. *Neurosci. Res.* 104, 80–87. doi: 10.1016/j.neures.2015.12.008

Huttenlocher, P. R., Courten, C. D., Garey, L. J., and Loos, H. V. D. (1982). Synaptogenesis in human visual cortex — evidence for synapse elimination during normal development. *Neurosci. Lett.* 33, 247–252. doi: 10.1016/0304-3940(82)90379-2

Ivanenko, Y. P., Poppele, R. E., and Lacquaniti, F. (2004). Five basic muscle activation patterns account for muscle activity during human locomotion. *J. Physiol.* 556, 267–282. doi: 10.1113/jphysiol.2003.0 57174

Kargo, W. J., and Giszter, S. F. (2000). Rapid correction of aimed movements by summation of force-field primitives. *J. Neurosci.* 20, 409–426. doi: 10.1523/ JNEUROSCI.20-01-00409.2000

Kargo, W. J., and Giszter, S. F. (2008). Individual premotor drive pulses, not time-varying synergies, are the units of adjustment for limb trajectories constructed in spinal cord. *J. Neurosci.* 28, 2409–2425. doi: 10.1523/JNEUROSCI.3229-07. 2008

Kargo, W. J., Ramakrishnan, A., Hart, C. B., Rome, L. C., and Giszter, S. F. (2010). A simple experimentally based model using proprioceptive regulation of motor primitives captures adjusted trajectory formation in spinal frogs. *J. Neurophysiol.* 103, 573–590. doi: 10.1152/jn.01054.2007

Kaufman, M. T., Churchland, M. M., Ryu, S. I., and Shenoy, K. V. (2014). Cortical activity in the null space: permitting preparation without movement. *Nat. Neurosci.* 17, 440–448.

Kielar, A., Deschamps, T., Chu, R. K., Jokel, R., Khatamian, Y. B., Chen, J. J., et al. (2016). Identifying Dysfunctional Cortex: Dissociable Effects of Stroke and

Aging on Resting State Dynamics in MEG and fMRI. *Front. Aging Neurosci.* 8:22–32. doi: 10.3389/fnagi.2016.00040

Krouchev, N., Kalaska, J. F., and Drew, T. (2006). Sequential activation of muscle synergies during locomotion in the intact cat as revealed by cluster analysis and direct decomposition. *J. Neurophysiol.* 96, 1991–2010. doi: 10.1152/jn.00241. 2006

Kutch, J. J., and Valero-Cuevas, F. J. (2012). Challenges and new approaches to proving the existence of muscle synergies of neural origin. *PLoS Comput. Biol.* 8:e1002434. doi: 10.1371/journal.pcbi.1002434

Levine, A. J., Hinckley, C. A., Hilde, K. L., Driscoll, S. P., Poon, T. H., Montgomery, J. M., et al. (2014). Identification of a cellular node for motor control pathways. *Nat. Neurosci.* 17, 586–593. doi: 10.1038/nn.3675

Lillicrap, T. P., and Scott, S. H. (2013). Preference distributions of primary motor cortex neurons reflect control solutions optimized for limb biomechanics. *Neuron* 77, 168–179. doi: 10.1016/j.neuron.2012.10.041

Mendelson, B., and Frank, E. (1991). Specific monosynaptic sensory-motor connections form in the absence of patterned neural activity and motoneuronal cell death. *J. Neurosci.* 11, 1390–1403. doi: 10.1523/JNEUROSCI.11-05-01390. 1991

Morasso, P. (1981). Spatial control of arm movements. *Exp. Brain Res.* 42, 223–227.

Overduin, S. A., d'Avella, A., Roh, J., and Bizzi, E. (2008). Modulation of muscle synergy recruitment in primate grasping. *J. Neurosci.* 28, 880–892. doi: 10.1523/ JNEUROSCI.2869-07.2008

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.

Santello, M., Flanders, M., and Soechting, J. F. (1998). Postural hand synergies for tool use. *J. Neurosci.* 18, 10105–10115. doi: 10.1523/JNEUROSCI.18-23-10105. 1998

Takei, T., Confais, J., Tomatsu, S., Oya, T., and Seki, K. (2017). Neural basis for hand muscle synergies in the primate spinal cord. *Proc. Natl. Acad. Sci. U.S.A.* 114, 8643–8648. doi: 10.1073/pnas.1704328114

Takei, T., and Seki, K. (2010). Spinal interneurons facilitate coactivation of hand muscles during a precision grip task in monkeys. *J. Neurosci.* 30, 17041–17050. doi: 10.1523/JNEUROSCI.4297-10.2010

Takei, T., and Seki, K. (2013). Synaptic and functional linkages between spinal premotor interneurons and hand-muscle activity during precision grip. *Front. Comput. Neurosci.* 7:40. doi: 10.3389/fncom.2013.00040

Ting, L. H., and McKay, J. L. (2007). Neuromechanics of muscle synergies for posture and movement. *Curr. Opin. Neurobiol.* 17, 622–628. doi: 10.1016/j. conb.2008.01.002

Torres-Oviedo, G., Macpherson, J. M., and Ting, L. H. (2006). Muscle synergy organization is robust across a variety of postural perturbations. *J. Neurophysiol.* 96, 1530–1546. doi: 10.1152/jn.00810.2005

Tresch, M. C., and Jarc, A. (2009). The case for and against muscle synergies. *Curr. Opin. Neurobiol.* 19, 601–607. doi: 10.1016/j.conb.2009.09.002

Tresch, M. C., Saltiel, P., and Bizzi, E. (1999). The construction of movement by the spinal cord. *Nat. Neurosci.* 2, 162–167. doi: 10.1038/5721

Yan, Y., Goodman, J. M., Moore, D. D., Solla, S. A., and Bensmaia, S. J. (2020). Unexpected complexity of everyday manual behaviors. *Nat. Commun.* 11:3564. doi: 10.1038/s41467-020-17404-0

Yang, Q., Logan, D., and Giszter, S. F. (2019). Motor primitives are determined in early development and are then robustly conserved into adulthood. *Proc. Natl. Acad. Sci. U.S.A.* 116:201821455. doi: 10.1073/pnas.1821455116